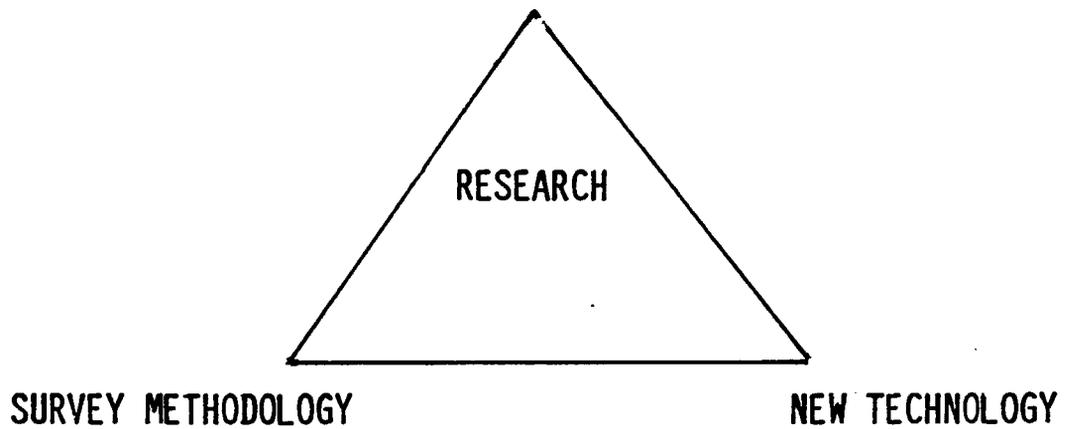


RESEARCH GOALS AND ACTIVITIES

STATISTICAL THEORY AND PRACTICE



STATISTICAL RESEARCH DIVISION
STATISTICAL REPORTING SERVICE
U.S. DEPARTMENT OF AGRICULTURE

MARCH 1986

CONTENTS

	<u>PAGE</u>
INTRODUCTION	1
OVERVIEW AND MISSION OF REMOTE SENSING RESEARCH	2
OVERVIEW OF SAMPLING FRAMES AND SURVEY RESEARCH	11
OVERVIEW OF OBJECTIVE YIELD RESEARCH	24

1970

DEPARTMENT OF AGRICULTURE

RESEARCH REPORT

RESEARCH ON COLLECTIVE FARMING

INTRODUCTION

The Statistical Reporting Service's programs are organized in the following major areas: Crop and livestock estimates, statistical research and service, and work performed for others. Research is conducted to improve the Statistical methods and techniques used to produce agricultural statistics. This research is done in support of the SRS long-range program for improving the accuracy of crop and livestock estimates at minimum cost and is directed toward better sampling, yield forecasting, and survey techniques.

The purpose of this report is to provide an overview of the research presently underway and research proposed for the future.

An issue facing any research organization is to determine where to focus its research efforts. The statistical research program in the Statistical Reporting Service has been subjected to a very thorough review during the last year to determine where research needs to be directed and what areas need to be emphasized. The research program can be described as involving three major activities:

- (1) Research on statistical methodology - This involves the development of new sampling and estimation procedures as well as a continual review of current methodology being used. Some examples of the type of research involving statistical methodology include the development of imputation procedures to estimate for missing data and refusals. Continual work is underway to improve the crop yield forecast models. The implementation of the Integrated Survey Program is raising many issues concerning optimum sample design for multi-purpose multiple frame surveys as well as improved multiple frame estimators. Recent task force reports such as the Crop Reporting Board Standards report strongly emphasize the need for more data analysis and the development of composite estimators. Other research includes estimation using remotely sensed data.
- (2) Research on new technology - Technology research involves developing the use of remote sensing technology and computer assisted telephone interviewing methodology. Other research involves evaluating the use of supercomputers to handle large data sets created by the remote sensing program. Other technology involves development of procedures to video digitize segments and the use of statistical graphics procedures. The evaluation and implementation of the new technology is guided by appropriate use of statistical theory and practice.
- (3) Research on improving survey procedures - It is becoming widely known that merely changing the order of questions or the wording of a question will change the level of an estimate. The evaluation of questionnaire design concepts is an important research topic. Considerable research is underway to evaluate the use of historic data in an interviewing situation. Other research efforts focus on process quality control which involves the evaluation of all steps in a sample and survey process to ensure procedures were defined and correctly followed.

The following sections outline the research program. The research is outlined by defining goals and a workplan for each research project.

Comments, questions and suggestions about the research goals will be appreciated.

OVERVIEW AND MISSION OF REMOTE SENSING RESEARCH

The primary mission of the Remote Sensing Research program is to 1) Explore alternative uses of satellite data for application to Agency programs; 2) Develop operational procedures for those applications; 3) Evaluate new methodology, and; 4) Evaluate data from alternative satellites and sensors.

SRS became involved in this area of research in 1972 when Landsat I was launched. The SRS approach for using satellite data has been 1) Remote sensing is simply another method of data collection; 2) Remote sensing can supplement the existing SRS data collection system but never completely replace it; 3) Data collection from satellites must be integrated with existing ground data surveys through rigorous statistical methodology, and; 4) Resource effective techniques must be developed for successful integration of this technology.

SRS has made good progress in this research area over the years. We have learned how to use satellite data for estimating acreage of major crops and made some progress in specialty crop estimation. However, these techniques are not yet cost effective. There is also concern over the consistent level difference between remote sensing estimates and the June Enumerative Survey. The amount of time required to make these estimates and the people resources needed also must be reduced further. Satellite images are used as part of the operational area frame construction process as a result of this research. Initial efforts to use satellite data for yield estimation were disappointing. However, some recent research in yield estimation with satellite data looks promising.

FIELD LEVEL EDIT OF AREA FRAME DATA

- Goal:** Develop edit procedures which allow data to be captured and edited at the field level during the operational June Enumerative Survey. (JES)
- Background:** Field level JES data are required as "ground truth" for input to remote sensing procedures. These data are presently obtained by capturing field level data records keyed for the JES and re-editing them for consistency and one to one correspondence between the edited field level data and the fields on the photo. The field level edit is currently done after the JES which requires large expenditures of travel monies. A JES field level edit will save many resources. Other advantages to SRS include improved quality control of JES crop data and automated selection of objective yield samples.
- Work Plan:** Plans are being developed by Systems Branch to create a SAS edit for Enumerative Surveys. Remote Sensing Branch will work with Systems Branch to ensure that the new edit will satisfy "ground truth" data needs for remote sensing applications. The new systems may be available for use on remote sensing project states in 1987. Eldon Thiessen will be the Remote Sensing Branch coordinator.

REMOTE SENSING INDICATIONS

- Goal:** Provide Remote Sensing indications for major crops in Arkansas, Colorado, Illinois, Indiana, Iowa, Kansas, Missouri and Oklahoma.
- Background:** Crop acreage estimation for major crops began as part of the AgRISTARS program in 1980. Since 1980 the project has grown from two states to eight for the 1986 crop year.
- Work Plan:** Estimates will be provided for Arkansas, Colorado, Illinois, Indiana, Iowa, Kansas, Missouri and Oklahoma.
- Crop coverage will be as follows:

State	Crop					
	Winter Wheat	Corn	Cotton	Rice	Sorghum	Soybeans
Arkansas	X		X	X	X	X
Colorado	X					
Illinois		X				X
Indiana		X				X
Iowa		X				X
Kansas	X					
Missouri	X	X	X	X	X	X
Oklahoma	X					
Percentage of U.S. Acreage	46	41	6	45	14	49

Estimates will be provided to meet Crop Reporting Board year end due dates. Eldon Thiessen will coordinate these activities.

COUNTY ESTIMATES

- Goal:** Provide County and crop reporting district indications for major crops from remote sensing techniques.
- Background:** County and crop reporting district estimates of major crops were done in all states in the DCLC project in 1985. This program will be continued for 1986 with some software modification to improve processing.
- Work Plan:** County and crop reporting district estimates for major crops will be provided to Arkansas, Colorado, Illinois, Indiana, Iowa, Kansas, Missouri and Oklahoma for the 1986 crop year. Indications will be provided in three formats; in hard copy printout, in a file that can be imported into Lotus 1-2-3, and a file of card image records that can be input to the agencies county estimation software. Eldon Thiessen will be the coordinator for the Remote Sensing Branch county estimate work.

DOCUMENTATION

- Goal:** Create user documentation and a training module outlining procedures for making estimates from Landsat and JES data.
- Background:** The current procedures for using remote sensing techniques to produce acreage estimates are only partially documented. A training module for new statisticians is also needed. This documentation will make it easier to train new people and make it possible to involve other people in SRS in the estimating process without experiencing mistakes that could undermine the program. All branch processing for small scale estimation work is being moved to the Martin Marietta Data Systems center used by the agency for small scale estimation work for all data processing.
- Work Plans:** User documentation will be created by Branch members as part of the conversion effort. Documentation should be completed by March 1987. Documentation of PEDITOR programs is being written as the programs are converted to PASCAL. Eldon Thiessen will serve as project leader for user documentation and Richard Sigman will serve as project leader for PEDITOR program documentation.

REFINE SOFTWARE

- Goal:** Make refinements to PEDITOR software modules and processing procedures to streamline processing of Landsat data for crop acreage estimation and increase the use of batch processing. (PEDITOR is a set of software that SRS has designed to process Landsat data.)
- Background:** The current PEDITOR programs have been developed as stand alone modules to process Landsat and JES data for acreage estimates. The system is highly flexible; however, the process requires considerable intervention of experienced personnel in order to complete the analysis. These refinements will make the system easier to use, provide greater safeguards against mistakes, and reduce costs and analyst time. The software will be more suitable for use in an "operational" environment. Refinements will also be made to job streams and processing procedures to promote efficiency.
- Work Plans:** EDITOR is a collection of computer programs used by SRS for processing Landsat data. PEDITOR is a derivative of EDITOR that can be transported to several different computer systems. The development of PEDITOR has involved rewriting various EDITOR programs in Pascal. This work is being performed by both SRS and NASA programmers and will be completed by the fall of 1985. Some of the identified refinements are being incorporated during the rewriting process --especially in programs being written by SRS. The remaining refinements should be completed by mid- to late- 1986.
- Many of the other improvements to processing procedures will be addressed as we convert processing from BBN and NASA/Arnes to the new computer centers in the fall of 1985 and the spring of 1986. Martin Ozga will be the technical coordinator of the project.

USE OF MULTITEMPORAL SATELLITE DATA

- Goal:** Expand the use of multitemporal satellite data to estimate corn, soybeans and wheat acreage.
- Background:** Most of the estimates that SRS has produced using satellite data have been made using a single date of imagery combined with JES ground truth data. For the 1985 season multitemporal data was used to estimate winter wheat in Oklahoma and winter wheat and spring planted crops in Arkansas and Missouri. A fall 1984 scene was combined with a spring 1985 scene for Oklahoma wheat estimates. A combination of spring and summer scenes was used for estimates in Arkansas and Missouri.
- Workplan:** The use of multitemporal coverage will be expanded to more states as processing procedures are improved. This will provide additional crop coverage without adding new States to the program.

MAP PRODUCTS FROM REMOTE SENSING

- Goal:** Evaluate the use of remote sensing for generation of map products, specialty crop estimation and the use of microcomputers for processing satellite data.
- Background:** SRS is currently involved in a cooperative project with the University of California at Berkeley, NASA/Ames and the California Department of Water Resources (DWR) to develop procedures and software to estimate specialty crops and provide map products for California. The project developed out of the needs of DWR to map crop acreages to estimate demands for irrigation water and the SRS mission to provide small area estimates for many of the same crops.
- Work Plan:** An operational test will be conducted in 1985 using data from JES segments and transect data collected by DWR from ground observations. These data will be combined with three dates of satellite imagery to produce state and county estimates for major and specialty crops and map products for use by DWR and the California SSO. Estimates will be completed by late December for use by the SSO and Crop Reporting Board. County estimates and map products will be completed in 1986. Most of the data will be processed on a microcomputer; however, a supercomputer will be used for full frame processing.

Results from this project will help evaluate the use of satellite data for this application, the ability of a microcomputer to process satellite data, the feasibility of distributive processing and the feasibility of a cooperative effort to collect ground truth data. Richard Sigman will be the technical coordinator.

REGRESSION ESTIMATOR

Goal: Find and eliminate the cause of consistent large downward bias of the remote sensing regression estimator when compared to the direct expansion of the JES.

Background: Since 1978, we have produced 40 estimates of corn, soybeans and winter wheat. In thirty-five cases, the regression estimate from remote sensing has been below the JES direct expansion, while the regression estimate has been above the JES only five times. Simulation studies conducted for SRS by Lockheed argue that the two estimators should be estimating the same level within a one percent relative difference. Several possible explanations are: 1) Lockheed simulation results have under-estimated the bias; 2) Expansion errors; 3) Aggregation errors when combining analysis district estimates to a state total; 4) Classifier overfitting due to the use of JES data for both training the classifier and estimating regression parameters; 5) Classifier overfitting due to "too much looking at the data" while developing the classifier.

Work Plan: The Remote Sensing Branch will continue to investigate possible errors in expansion or aggregation. However, these areas have been looked at before. A classifier training study will be conducted in Iowa and Missouri in the summer of 1985 to investigate the issue of non-independent training data and overfitting the classifier. A time-table for activities is as follows:

1. Collect ground data in Iowa and Missouri July-August 1985
2. Field level edit September 1985
3. Digitize segments September-October 1985
4. Data analysis November-April 1986
5. Report Completed August 1986

Richard Sigman will be the technical coordinator.

EVALUATE SUPERCOMPUTERS

Goal: Compare the CRAY-XMP, CYBER-205 and the Massively Parallel Processor supercomputers for maximum likelihood classification of Landsat data.

Background: The Massively Parallel Processor (MPP) is a supercomputer at the Goddard Space Flight Center which consists of many processors operating in parallel as opposed to the pipeline processors of the CRAY and CYBER computers. The processing of large amounts of data in a timely and economical fashion with minimal manual intervention is important if remote sensing is going to be used by SRS. This project will allow us to evaluate three supercomputers for this application.

Work Plan: A research proposal has been sent to Goddard Space Flight Center to test this application. If the proposal is accepted, work will be completed in one year. There will be no charge for use of the MPP. CRAY and CYBER processing will be completed at NASA/Ames under existing agreements. Martin Ozga will be the technical coordinator for this project.

REMOTE SENSING FOR AREA FRAME

Goal: Explore the use of digital satellite data and a microcomputer based image processing system for area frame construction and updating.

Background: A project to explore the use of digital satellite data for area frame construction was conducted by the Fairfax Stratification Unit and Greg Burns at NASA/NSTL. This project would be a follow on to refine workable techniques and explore new applications. The use of digital satellite data and an image processing system for area frame construction has the following potential benefits: 1) Frame materials would be placed in a geographically referenced data base for easy use. A series of 1:100,000 digital maps are being developed by USGS for the 1990 census. 2) Count unit identification and digitization could be combined into one process. 3) Stratification materials could be referenced to the same scale and overlaid to avoid manual transfer of boundaries. 4) Materials can be modified to enhance features needed for stratification. 5) Allow frame updating rather than complete reconstruction. 6) A digital data base will improve our ability to create crop specific stratification. 7) Map products can be created for quality control. 8) All area frame processing can be done on the same machine which will reduce costs and time required for frame construction and maintenance.

Workplan: Stratification research will be coordinated with the Fairfax Sampling Unit. Initial work will be done in New York to construct a frame to estimate acreage of orchards and vineyards. This project will provide a measure of incompleteness in the New York Orchard and Vineyard Survey which will be conducted from a list sample in the fall of 1985 and the spring of 1986. The New York area frame is already digitized which will make the project easier. We also have a cooperative agreement with Cornell University to develop procedures to use satellite data to identify orchards and vineyards. We will develop a classified image of orchards, overlay the current count unit boundaries, assign count units to a new strata and select a sample of segments. Segments will be enumerated to provide an estimate of nonoverlap from the orchard list sample. A regression estimate of fruit and vineyard acreage will be developed. New York will pay for data collection and area frame materials. Work will be completed in late 1986.

Follow on research will be conducted to design procedures applicable for general area frame construction and maintenance. These activities will begin in FY1986. Marty Holko will be the technical coordinator for this project.

IMPROVED VIDEO DIGITIZING PROCEDURES

- Goal:** Investigate the use of a microcomputer and color video camera for digitizing and labeling JES segment and field boundaries for remote sensing projects. This procedure may also have some application for digitizing boundaries for area frame construction. Commercially available software for tablet digitizing will also be investigated.
- Background:** The current video digitizing procedures were designed using a mini-computer and a black and white T.V. camera. This equipment was "state of art" several years ago but does not lend itself to decentralization of video digitizing. The microcomputer based system will also be cheaper than the current system. The color camera may also make the labeling process easier. The use of commercially available software for tablet digitizing would have uses in Remote Sensing Branch and the Fairfax Sampling Unit.
- Work Plan:** The use of this equipment and possible applications to area frame construction will be explored in 1986. We hope to have procedures developed by the end of 1986. The project will require about \$10,000 for a microcomputer, color TV camera and other hardware. Richard Sigman will be the technical coordinator.

DATA FROM ALTERNATIVE SATELLITES

- Goal:** Evaluate data from the French SPOT satellite data for SRS needs.
- Background:** SRS is involved in a bilateral remote sensing agreement with the French. SRS will provide our software for making regression estimates from satellite and ground truth data to the French in exchange for data from the French SPOT satellite. Sites over Iowa and Kansas have been chosen for the study.
- Work Plan:** The SPOT satellite was launched in February of 1986. We are hoping to acquire data in April or May 1986 over Kansas which will be compared to results obtained from MSS data from the U.S. Landsat satellite for estimating winter wheat. Data will be acquired in July or August over Iowa for use in estimating corn and soybean acreage. Data analysis should be completed by mid 1987. Richard Sigman will be the technical coordinator for the project.

CROP CONDITION ASSESSMENT

- Goal:** Investigate the use of MSS and AVHRR satellite data and weather data for crop condition assessment.
- Background:** SRS signed a Memorandum of Understanding with the Foreign Agricultural Service, Agricultural Stabilization & Conservation Service and Agricultural Research Service in mid 1985 to participate in joint research to explore the use of satellite and weather data for crop condition assessment. These data are currently available over part of the U.S. because of an ASCS program to monitor disasters and growing conditions by means of visual inspection of satellite scenes. FAS uses similar data to assess foreign crop production potential. We want to determine if quantifiable relationships can be developed from this information.
- Workplan:** We have begun working with FAS to determine the type and frequency of data available, the amount of U.S. coverage available and the data manipulation capability of the FAS image analysis system. We have also reviewed weather analysis procedures used by meteorologists at the Joint Agriculture Weather Facility. We plan to evaluate possible information products such as maps of floods, freezes and winterkill damage that could be produced to improve the usefulness of the Weekly Weather Crop Report. Statistical relationships between SRS data series such as crop condition, yield, plant counts, etc. and vegetative indexes, soil moisture or plant stress and evaporation will also be explored. Wendell Wilson will be the technical coordinator of this project.

USE OF THEMATIC MAPPER

- Goal:** Investigate the use of Thematic Mapper satellite data for acreage and yield estimation.
- Background:** A 1982 scene of TM data in Iowa was investigated for possible acreage and yield estimation in 1984. This project showed a strong correlation between farmer reported soybeans yield and satellite data values. The TM data has much smaller resolution and seven bands of reflectance readings which may also improve crop discrimination.
- Work Plan:** An analysis of 1985 TM scenes in Missouri and Iowa will be done. These were chosen for the study because crop data will be available for additional sample replicates for these TM scenes as a result of the 1985 classifier study. An Indiana TM scene may be used to examine yield relationships. This was chosen because of the additional soybean objective yield data that will be available from the 1985 validation project. Analysis of these data sets will be completed in the future as staff permits. Richard Sigman will be the technical coordinator of the projects.

EVALUATION OF STANDARD SRS HARDWARE

- Goal:** Explore the feasibility of using a Unix based system such as Inforex or an IBM PC for calibrating and digitizing JES segments and Landsat scene registration.
- Background:** Programs for digitizing and registration were developed for a North Star microcomputer and a PDP 11/44 minicomputer. While this equipment is functional, it is becoming out of date and there are only a few SRS personnel that are familiar with this equipment. The use of "standard" SRS hardware for remote sensing would provide the Remote Sensing Branch with many SRS personnel that are familiar with the equipment; allow utilization of remote sensing equipment for other SRS applications during periods when it was not needed for remote sensing; and provide up to date equipment for remote sensing.
- Work Plan:** Work on this project will begin in late 1986, after the conversion of small scale processing to MMDS. Richard Sigman will be the technical coordinator.

OVERVIEW OF SAMPLING FRAMES AND SURVEY RESEARCH

The immediate and long-term goals for Sampling Frame and Survey Research can be categorized into two general areas: (1) area frame construction, management, and research and (2) information collection research.

Area frame construction involves the updating of current frames for SRS use in the Agency operational survey programs. Efficient updating and maintenance of these frames requires several manual and automated processes. Improvements in all phases of area frame construction and maintenance are continually being investigated.

The efficient collection and handling of survey response information will be highlighted in our research efforts. Adjusting for missing information, evaluating respondent effects and proper use of prior survey data will be investigated. Computer assisted telephone interviewing procedures are being phased into the operational program to standardize telephone procedures and improve data quality.

Statistical procedures for analyzing survey results via graphics and utilizing composite estimation by the Crop Reporting Board will be developed.

COMPOSITE ESTIMATION

Goal: Develop a statistical procedure that can be used to combine survey estimates (nonprobability and probability) to arrive at an overall estimate with some measure of precision. This procedure must be statistically defensible and repeatable. Another dimension of this study would take the national estimate adopted by the Crop Reporting Board and statistically set the regional and state level estimates.

Background: Procedures for evaluating estimates (nonprobability and probability) are currently subjective. Staff members utilize survey indications and check data from varied sources at review time and subjectively establish a weight for each survey estimate to determine the final estimate. Composite estimation has been used by numerous statisticians and was anticipated by Agency statisticians (Houseman 1971). In a recent Agency publication (Crop Reporting Board Standards), the need for such a procedure was voiced. Lynn Kuo, American Statistical Association Research Fellow, 1985, has recently completed an initial investigation of a composite estimation model used to combine four estimators. The composite estimator is derived by minimizing a quadratic function subject to linear constraints. The variance and mean squared error are evaluated by the jackknife method.

Work Plan: Based on the composite estimation methodology (Kuo 1986), further research is necessary to expand the technique to generalized use. This includes use for other commodities, other States, other preliminary estimators and to second stage sampling applications. Numerical results will be expanded to commodities other than hogs. Development of an algorithm to calculate State level estimates from the official National level estimate is also a priority. Integrating the composite estimation into the current crop reporting board activities will be another objective. Further research on variance evaluation is needed.

Brian Carney will be the technical coordinator.

SMALL AREA ESTIMATION

Goal: The Agency goal is to implement a defensible and repeatable procedure to estimate for small areas (counties) within a State. The short-term goal is to make this procedure operational in North Carolina and continue testing in at least one additional State. The long-term goal is to develop a probabilistic procedure for general use in the operational program in estimating county level data in all States.

Background: For several years, the Agency has been establishing county level estimates for crops and livestock using subjective procedures. Techniques for constructing county level estimates which are statistically valid, defensible, and repeatable have been investigated in recent years. There are three background reports published by the Statistical Research Division, "The Development of County Estimates in North Carolina" Barry Ford, 1981; "Combining Historical and Current Data to Make District and County Estimates for North Carolina," Barry Ford, Doug Bond, and Nancy Carter, 1983^{1/}; and "An Evaluation of Categorical Data Analysis Methodology for County Estimates in North Carolina," Nancy Carter and Doug Bond, 1985. The latest report by Nancy Carter provides an evaluation of three Categorical Data Analysis (CDA) estimators to derive county-level harvested acreage estimates for several crops in North Carolina. These three estimators, Case 1 (full association structure), Case 2 (partial association structure), and Case 3 (iterative proportional fitting), were each evaluated using 1978 North Carolina Census of Agriculture data and the 1981, 1982, and 1983 A&P survey data.

^{1/} Nancy Carter was a participant in the Research Institute Fellowship between SRS and the American Statistical Association.

Work Plan:
FY '86

- (1) Continue evaluation of Categorical Data Analysis (CDA) Methodology for acres harvested variables with 1985 North Carolina county level estimates.
- (2) Evaluate the CDA generated estimates against 1982 Census data in North Carolina.
- (3) Consider other synthetic estimators and other sources of information for benchmark comparison.
- (4) Begin evaluation of methodology in a second State (California) and also use 1982 Census data for benchmarking.
- (5) Establish procedures to handle double cropping.
- (6) Consider methods to handle data disclosure issues and to streamline obtaining of data from Census.

FY '87

- (1) Expand the evaluation to include production data and livestock data variables.
- (2) Expand the evaluation time period to include 1982-1985 A&P survey data and 1978 and 1982 Census data. (California will use 1982 Census)
- (3) Evaluate remote sensing information in California.
- (4) Begin evaluation in a third probability Acreage and Production Survey State.
- (5) Consider alternative CDA methods.
- (6) Plan to expand concept into the operational program.

Doug Kleweno will coordinate the Agency effort. Work will continue under a Cooperative Agreement with California State University-Chico. Nancy Carter represents the Cooperator.

COMPUTER ASSISTED DATA COLLECTION

Goal: Implement a plan of collecting data using technology identified as computer assisted survey execution system (CASES). The goal is to have this technology operational for telephone interviewing in the Agency in several States during FY'86. The feasibility of portable data entry and other applications using the CASES software will be investigated.

Background: Initial research in this technology began in the late 1970's. In January 1981, a research agreement was signed with the University of California at Berkeley for development of the computer assisted telephone interviewing (CATI) software. From January 1982 through mid-1985, software testing in California and Nebraska (June 1984) has been underway. During the early test phase, several studies were undertaken to evaluate various aspects of CATI. These include: effect on level of estimates for hog and cattle surveys, enumerator characteristics, office procedures, completion rates, and response time. Reports are available from the Sampling Frames and Survey Research Branch.

Work Plan: FY '86

- (1) Develop and begin implementation of a plan to make CATI operational in the SSO's.
- (2) Train transition team: CATI procedures, CASES software, "Q" and "C" programming.
- (3) Prepare Agency training and user materials.
- (4) Test CASES software after successful porting to Inforex SSO machines.
- (5) Conduct training school(s) and support States in their training and initial survey efforts.
- (6) Support existing States with Zilog microcomputers.
- (7) Develop instrumentation including documentation
- (8) Research:
Historical data usage with CASES
Automatic Scheduler Efficiency Test
MS-DOS Efficiency Test
- (9) Continue CASES software enhancements
- (10) Determine position for "caretaker" of CASES software.

FY '87

All Inforex States should be "fully" operational with CATI being used for selected telephone surveys. Continue expansion of the CATI program into additional States as CATI team support and Agency hardware permits. Activities include:

- (1) Develop additional survey instruments for all probability surveys using telephone enumeration. Include front end data edit and reported data capture.
 - (2) Conduct CATI training school and provide SSO support.
 - (3) Research:
Develop a cost model for CATI
Evaluate CATI enumerator efficiency based on experience
Self-administered questionnaire development
Investigate use of portable data entry devices
CASES software used in a local area network and database environment
 - (4) Develop specifications for omnibus RFP to include CASES in all SSO's
 - (5) Continue CASES software development
 - (6) Train additional Agency staff on "Q" and "C" programming
 - (7) Rewrite cooperative agreement for computer assisted software research and development
- Dick Coulter and Doug Kleweno will coordinate this effort.

USE OF HISTORICAL AND PREVIOUSLY REPORTED SURVEY DATA

Goal: Determine the need, method(s), and occasion(s) to use historical survey data during data collection. The impact of various methods on sampling and nonsampling errors and level of the estimate are major issues.

Background: The Agency uses historical survey data in many surveys to evaluate responses during the data collection process. Each State Statistical Office uses prior survey information in a variety of ways with no established rules. This affects the current survey results in an unknown manner. Prior survey data are used as a measure of size on the LSF so it is available as control information on the label provided for the sampled unit. The data are also available for use when survey probing is necessary, for data imputation and when computer editing is carried out. There is also the concern for data security and who reported previous information for the sampled unit.

Work Plan: Steps are currently underway to complete a background search on the subject, including a literature review, and to evaluate proper use of historical data. A formal study was recently conducted in California. The California test used a computer assisted survey instrument to evaluate different treatments of historical data on the 1985 Acreage and Production Survey. Respondents reporting data on an earlier survey (June Acreage Survey) were contacted. Four treatment groups were used: 1) no previous data provided, 2) previous data available but not given to respondent, 3) previous data provided to respondent to update and change as part of question, and 4) previous data were available for verification and edit check only. A randomized complete block design was used. A preliminary report was given at a research forum in February. A final report will be forthcoming.

In April 1986, a followup study concerning grain stocks will be conducted in California, Nebraska, and Georgia using historical data. Two treatments are planned. Data will be collected by telephone for the entire sample using CATI.

Brad Pafford and Doug Kleweno are technically responsible for this research effort.

BRIDGE SURVEY

- Goal:** Provide a measure of the effect of changing to a "first of the month" reference date and collecting data during the 15 workdays on or after the first.
- Background:** A task force, created to study the Agency's survey reference dates for specific commodities, recommended changing to common survey reference dates across commodities. This was intended to coordinate data collection activities into a quarterly time frame with similar reference periods for all commodities. This requires measuring potential change in level of inventory when asking the respondent to report as of December 1, for example, instead of at the time of the interview (data collection between November 21 and December 7).
- Work Plan:** Evaluate potential memory bias associated with asking questions related back to the first of the month. Hogs, cattle, and grain stocks questionnaires are also designed to provide comparable data to test the bridge and operational survey total inventories for significant level differences between the current operational and proposed reporting date. Bridge surveys will be conducted for a June 1 and December 1, 1986 reference date in several States. Statistical tests will be made for list surveys. Area samples will be used for questionnaire evaluation only.
- Brad Pafford will be responsible for analysis of data.

RESPONDENT AND DATA COLLECTION METHOD EFFECTS

- Goal:** Examine the incidence of respondent and data collection effects in Agency data and determine whether changes in technique will reduce this incidence.
- Background:** There is considerable evidence in the sampling literature of biases in survey results due to changes in the method of data collection such as from personal interview to telephone interview, varying the number of attempts to contact the respondent (hence indirectly the timing of the contact), and changes in respondent (from operator to spouse or other knowledgeable individual). This project will attempt to address such issues in the context of the farm operator as the preferred respondent in Agency surveys.
- Work Plan:** Data from the September 1985 survey in nine ISP test States were examined to detect differences in quality due to the respondent (operator, spouse, or other knowledgeable individual) and also to detect differences due to method of data collection (mail, telephone or personal interview). Questions to capture the number of attempts made and the time of completion for telephone interviews have been added to the Crop Integrated Survey Program (CRISP) instruments for five States. Information from this study may help in suggesting better scheduling of telephone contact attempts in order to maximize the probability of reaching the farm operator rather than the spouse or other individuals. Data collected from the 1985 California Acreage and Production Survey and 1986 April Grain Stocks Survey (Georgia, Nebraska and California) will also be used in this evaluation.

Bill Warde, a 1985-1986 American Statistical Association Research Fellow, will be the investigator.

DATA ANALYSES IN CROP REPORTING BOARD (CRB) PROCEDURES

Goal: Develop sound statistical procedures which can be used in a "real" time environment to evaluate survey estimates and raw data in setting National estimates during CRB activities.

Background: Analysis of survey information plays a major role in the current estimation procedures of the Agency. Analyses packages (including ADP programs) have been developed for SSO use in analyzing survey information. However, once analyzed at the State level, procedures for technical analysis at the headquarter's level are currently limited to subjective manual procedures. Data are not available in machine readable form to permit any analysis at the aggregate level to evaluate such things as influential data points and imputation methods.

Work Plan: Specific project objectives include:

- (1) Develop graphics and other computer displays for data review by CRB members and commodity statisticians. This includes the ability to identify atypical data points, etc. and determine influence on final estimate in a "real" time mode of processing.
- (2) Develop graphics procedures for presentations and for publication of official survey estimates.
- (3) Develop procedures for automated transmission of information to the CRB.
- (4) Evaluate and recommend hardware and software to support CRB activities.

Brian Carney will be the technical leader.

NEW AREA FRAME DEVELOPMENT

- Goal:** Construct new area frames in Montana, Delaware and Maryland for use in the 1986 June Enumerative Survey (JES) and in Nevada, New Jersey and possibly Utah for the 1987 JES.
- Background:** Each year a new area frame is constructed in a specified number of States (usually 2 or 3). This is due to the fact that land use has changed considerably since the previous area frame was developed. Current plans call for developing new frames by 1990 for States whose frames were constructed prior to 1974. This includes Utah, Nevada, New Jersey, Missouri, West Virginia and the six New England States.
- Work Plan:** Paul Blackwood will be responsible for developing the work plan and providing technical leadership in the construction (stratification and digitization) of new area frames for the 1986 and 1987 JES.

NEW AREA FRAME ANALYSIS

- Goal:** Conduct area frame analysis for each new area frame using the Area Frame Analysis Package.
- Background:** A regular part of new area frame construction is to evaluate the new area frame its first year in operation using data from the JES. Methods Staff and the Sampling Frame Development Section jointly evaluate the new frames.
- Work Plan:** Jim Cotter and someone from Methods Staff will conduct the analysis of the new area frames in Maryland, Delaware and Montana and document the evaluation by October 1, 1986.

NEW AREA FRAME DOCUMENTATION

- Goal:** Provide documentation of unique features for each area frame such as land use strata definitions, segment sizes, sample allocation, quality control procedures, results provided from the area frame analysis package, differences from the previous frame, etc. This documentation will contain the materials (memos, analysis, etc.) used to determine the characteristics of the area frame for the State.
- Background:** Thorough documentation for new area frames began several years ago when documentation binders were prepared for Texas and Idaho. These two frames were put in operation for the 1982 JES. Since then, six new frames have been constructed (Nebraska, Florida, Arizona, Colorado, New Mexico and Wyoming) and recently the documentation folders for all of these States were completed.
- Work Plan:** Paul Blackwood will be responsible for completing the documentation for Maryland, Delaware and Montana by October 1, 1986.

JUNE ENUMERATIVE SURVEY SAMPLE ROTATION

- Goal:** Select the sample of land areas (segments) and prepare the area frame photos for use in the June Enumerative Survey (JES). Approximately 4,000 segments are processed each year. This includes segments for new frames, segments rotated into the sample and corrective actions on problem segments from the previous year.
- Background:** Each year approximately 20 percent of the sample is rotated for use in the JES in order to reduce respondent burden. A sample of frame units is selected at random, frame units are broken down into segments, a sample of segments selected, the necessary photography ordered and finally the photography is prepared for use in the JES.
- Work Plan:** Marlowe Schlegel will be responsible for scheduling the workload in the Sample Select and Pen and Ink units so that the aerial photography is provided to the SSO's before April 1 to conduct each JES.

AGRICULTURAL ECONOMIC SURVEYS (AES) SAMPLE ROTATION

- Goal:** Provide approximately 930 new segments for the 1986 Agricultural Economic Surveys (AES) and about 800 new segments for the 1987 AES.
- Background:** Each year, Methods Staff requests that segments be rotated into the AES sample to reduce respondent burden. In the past, roughly 500 to 1,100 segments are rotated into the AES each year depending upon the priorities and staffing in the Sampling Frame Development Section.
- Work Plan:** Marlowe Schlegel will be responsible for reviewing the sample rotation requested by Methods Staff, determining how many segments can be provided and coordinating the work to provide the rotation sample for the AES.

AREA FRAME MASTER

- Goal:** Transfer responsibility of the area frame master from Data Collection Branch to the Sampling Frame Development Section for the 1986 JES.
- Background:** The area frame master is updated each year to reflect segments rotated into and out of the JES. During recent years, Data Collection Branch was responsible for maintaining the area frame master. The decision was made at the 1986 JES specifications meeting in October 1985 to transfer responsibility of the area frame master to the Sampling Frame Development Section.
- Work Plan:** Marlowe Schlegel will be responsible for coordinating the transfer of activities and for handling the annual update for the JES. Jim Mergerson will be responsible for the statistical review of the updates to the area frame master as well as the current segments on the area frame master.

JES DESIGN AND METHODOLOGY REPORT

- Goal:** Prepare a detailed technical report (approximately 200 pages) on the design and methodology of the JES as called for by the Crop Reporting Board Procedures Task Force.
- Background:** A 14-chapter table of contents was prepared by Jack Nealon and is being critiqued by numerous employees in SRS.
- Work Plans:** Tentative plans call for various members of SRS writing this report. Jack Nealon will likely be responsible for several chapters in the report. Writing will begin when the table of contents are approved by the Division Director.

DOCUMENTATION OF AREA FRAME PROCEDURES

- Goal:** Prepare a staff report that describes in detail the operational procedures used to stratify, digitize, select the sample and prepare the area frame materials for the JES. This report will serve as a training tool in area frame construction and sampling.
- Background:** Currently, documentation is available for each of the four units in Fairfax, however, this documentation needs to be updated and expanded to fully document the operational procedures.
- Work Plans:** Paul Blackwood will be responsible for the documentation on area frame stratification and digitization. Marlowe Schlegel will be responsible for the documentation on sample selection and preparing the photographs for use in sample surveys. Documentation for each unit will be updated by April 1, 1986. Jack Nealon will be responsible for integrating the documentation from the four units into a staff report. A draft staff report will be available by October 1, 1986.

AREA FRAME DATA BASE MAINTENANCE

- Goal:** The area frame information package, loosely called the area frame data base, needs to be updated annually with current photography dates, survey and official estimates, etc. The update procedures need to be changed from a manual to an automated process.
- Background:** The area frame data base was developed to be depository of information on each area frame to be used to evaluate area frames in conjunction with the area frame analysis package. These two packages will be used in the years ahead to determine objectively which States need new area frames. During 1985, the output tables from the data base were improved so that they are user friendly.
- Work Plans:** Stan Mason is responsible for updating the area frame data base each year and developing automated procedures for the annual updates.

DATA BASE MANAGEMENT SYSTEM

- Goal:** Begin to convert systems such as the JES Tracking System and the Agricultural Stabilization and Conservation Service/U.S. Geological Survey (ASCS/USGS) budget system to a Data Base Management System (DBMS) to provide quicker computer response time, more efficient data management and easier access to data.
- Background:** The JES Tracking System is an automated survey management tool that provides information about the status of the sample select activities, photo orders, the number of segments digitized, etc. The present data management system for the JES Tracking System, although very useful, can probably be improved by employing a Data Base Management System. The ASCS/USGS budget system is now done manually.
- Work Plan:** Work on evaluating the use of the DBMS for the JES Tracking System began in June 1985. Pat Thomas will be responsible for evaluating the DBMS during FY '86 and determining if a conversion of the JES Tracking System to the DBMS should be made. The ASCS/USGS budget system has been automated on the DBMS and will be implemented during FY '86. All other budget accounts such as training, travel, ADCOM, Xerox and DHL were automated using spreadsheet software by Kathy Souther and will be implemented during FY '86.

RANGE SEGMENTS

- Goal:** Develop a standard and consistent set of guidelines for SSO's on how to determine segment boundaries for range segments.
- Background:** There has been concern that the procedures used by SSO's for point samples are somewhat subjective and not standard from State to State. To minimize potential nonsampling errors, an evaluation of the procedures will be done to see if the procedures can be made more uniform.
- Work Plan:** Marlowe Schlegel will have the responsibility to evaluate the current point sample procedures and recommend improvements. Recommendations will be documented by August 1, 1986.

TRAINING

- Goal:** Improve the tools used to provide training to Agency and non-Agency personnel.
- Background:** To date, the "wall tour" has been the primary tool for conducting training at Fairfax. Recently, an 8-minute video was made on area frame construction.
- Work Plan:** The technical report that documents area frame design procedures will be available to Agency and non-Agency personnel so that area frame construction and sampling concepts can be learned without having to physically visit the Fairfax office. This report will also serve as an excellent training tool for new cartographic employees in Fairfax. In addition, two slide presentations will be developed by Jack Nealon for use away from the Fairfax location, e.g. JES training schools. These presentations will be: (1) an overview of area frame construction and sampling and (2) detailed presentation of area frame construction and sampling. Plans are to have the slide presentation completed by April 1, 1986 so that they can be discussed at the 1986 JES National School.

ENHANCED GRAPHICS FOR DIGITIZING

- Goal:** Enhance the graphics capability for area frame digitization.
- Background:** The Branch received the Tektronics equipment from Houston during September and Brian Carney will be testing the graphics capability until about February 1. Preliminary testing indicates such capabilities will enhance the area frame digitizing quality control procedures.
- Work Plan:** Pat Thomas will work with the Digitizing Unit and Brian Carney to develop a plan to enhance the graphics capabilities for digitizing during FY '86.

CENSUS OF AGRICULTURE

- Goal:** Provide additional urban and ag-urban segments for the 1987 Census of Agriculture.
- Background:** Tentative plans call for the Sampling Frame Development Section to select additional segments to be used in the 1987 Census of Agriculture.
- Work Plan:** Marlowe Schlegel will be responsible for coordinating the sample selection activities for the Census of Agriculture during 1986.

AES RESEARCH

- Goal:** Investigate the use of the weighted segment estimator in lieu of the open segment estimator and the use of JES-sized segments in lieu of AES-sized segments for economics surveys.
- Background:** A report prepared by Rich Allen and presented to the PPC during 1985 called for research into the use of the weighted estimator and alternative segment sizes for the AES.
- Work Plans:** Mike Clark will be responsible for conducting the analysis for the AES research project. Data will be collected during February and March 1986 and a draft report will be completed by September 1, 1986.

OPTIMUM SEGMENT SIZE

- Goal:** Develop and test statistical procedures for determining the optimum segment sizes to be used for new area frames.
- Background:** Currently, segment sizes for a State are based mainly on practical considerations with little consideration for what size is optimum statistically. Therefore, there is a need to develop procedures to determine what segment sizes will provide the greatest efficiency in terms of cost and precision. Also, this research will identify special data collection needs for determining the optimum segment sizes for States scheduled to receive a new frame.
- Work Plans:** Charles Perry is responsible for this project. The theoretical framework has already been formulated and the analysis is underway using 1985 data from Illinois. A draft report is expected by May 1, 1986.

UNIFORM SEGMENT SIZE

- Goal:** Provide statistical guidance on the tolerance levels that are acceptable when delineating segments within frame units during the sample selection process.
- Background:** Currently, segment sizes are allowed to fluctuate within certain tolerances in order to insure good boundaries and to try to provide homogeneity among the segments with respect to the amount of cultivated land in a segment. It is not known if the operational tolerance levels are too relaxed or restrictive.
- Work Plans:** Jim Mergerson is responsible for this project. Most of the analysis has been completed and a draft report summarizing the results of the basic simulation study is expected prior to May 1, 1986.

ROTATION EFFECTS

- Goal:** Determine if there is a rotation bias effect in the 1985 ISP surveys in Illinois. If so, identify the causes of the bias.
- Background:** With the advent of the ISP, it is imperative to determine what affect, if any, rotation groups have on the survey estimates and what is causing the differences, e.g. data collection procedures, handling of "known zeros," office editing and imputation procedures, underreporting of data, handling of refusals in follow-on surveys.
- Work Plan:** Jim Mergerson is responsible for research related to evaluating the effects and causes of rotation bias using 1985 ISP data in Illinois from the June, September and October surveys. An interim report will be prepared and a final report will be available around May 1, 1986.

LANDSAT STRATIFICATION PROJECT

- Goal:** Evaluate the use of Landsat digital data for area frame stratification.
- Background:** A study was conducted in selected counties in Florida and Wyoming to make a preliminary evaluation of the use of Landsat digital data as a tool for area frame stratification.
- Work Plan:** The final stage of photo interpretation for this study was completed in September 1985. Mike Clark will be responsible for conducting the statistical analysis and documenting the findings of this study before April 1, 1986.

AUTOMATED AREA FRAME CONSTRUCTION

- Goal:** Conduct initial research to evaluate the feasibility of using Landsat digital data and digital line maps to stratify and digitize area frames using microcomputers. This research will focus on updating area frames rather than constructing area frames from scratch as is currently done.
- Background:** This research will be pursued jointly by the Remote Sensing Section and the Sampling Frame Development Section. Its purpose is to make use of current technologies, to the extent possible, to automate and streamline the area frame construction process.
- Work Plan:** The work plan for this project is discussed under the goals for the Remote Sensing Section. Marty Holko will be the technical leader for this project. Jack Nealon will be the contact person for Fairfax during the preliminary developmental stage over the next year.

ESTIMATION RESEARCH

Goal: Develop a unified system of statistical procedures for editing, outlier detection and resolution, and data imputation.

Background: The Task Force Report on Crop Reporting Board Procedures called for development of consistent, statistical procedures for the editing of survey data, detection and resolution of outliers and the handling of item and entire nonresponse. The system must preserve the raw, edited, outlier adjusted and imputed data so that the effects of editing, outliers and imputation can be measured for each survey.

Work Plan: Jim Cotter will be responsible for the imputation research. June, September and possibly January ISP data in Illinois will be used in the study. A draft report on the June and September analysis is expected by April 1, 1986.

Charles Perry will be responsible for coordinating and monitoring research on robust estimation being conducted jointly by the Sampling Frame Development Section and Oregon State University. This research will get underway early in FY '86.

Mike Clark will be responsible for work in the area of statistical editing. Work will begin after Mike Clark completes the AES Research and Landsat Stratification Project.

OVERVIEW OF OBJECTIVE YIELD RESEARCH

Yield Research has two initiatives -- process control and forecasting. Process control involves an evaluation of sources of error from survey design, estimation, and nonsampling errors. Forecasting research is two-fold, addressing problems in the operating program and evaluating alternative techniques to find more reliable forecasts.

PROCESS CONTROL METHODS FOR OBJECTIVE YIELD SURVEYS

Goal: Improve the performance of objective yield surveys while demonstrating the usefulness of modern quality control philosophies for the control of Agency survey processes.

Background: Quality, in Deming's words, is the "searching out of root causes of trouble, and root sources for improvement." Traditional acceptance sampling, in current Quality Control (QC) procedures, validation surveys and various nonsampling error studies, are often regarded as "finger pointing" and can be destructive by creating an atmosphere of "who is to blame." Modern QC techniques stress process control and the need for changing organizational philosophies about the nature of quality issues. Despite the years of research which have been spent on objective yield (OY) techniques, an atmosphere of general unhappiness with OY surveys is present in the Agency. The techniques of process control show promise for explaining the surveys in a manner in which the impact of changes can be assessed.

The Yield Assessment Section (YAS) has investigated a number of OY survey topics and made proposed improvements. Examples of recent work include the use of fixed versus stepwise models for yield forecasting, plot location bias studies, plant handling and enumerator fatigue bias, the use of maturity categories in modeling, time lags between enumerator and farmer harvest, and problems with several estimators. Results of the studies concerning statistical techniques are usually accepted very slowly because of the difficulty in explaining their statistical impact, while changes which simplify operations may be adopted more easily.

Work Plans: The techniques of process control will be adapted to the OY survey process. This first involves the development of a "blueprint" of the "construction" of an OY survey estimate. This "blueprint" or process model will show the various steps in the survey process, the uses of the data collected, and information about the known or hypothesized contribution to sampling and nonsampling errors (an error profile).

Leadership of the overall model will be by Ron Fecso, with an anticipated rough draft date of spring 1986. Field level steps were written by Vince Matthews, and published in November 1985. Field selection processes have been studied cooperatively by Ron Fecso and Iowa State University personnel; results have been distributed within the agency. Based on the feedback, further studies (resampling plans and model errors) are in progress.

This work will provide a structure in which the potential improvements from proposed research and the impact of proposed survey changes can be assessed.

Under the process model it should become apparent that OY research generally is of three types—

Nonsampling error studies,
Survey design and estimation, and
Forecasting.

Goals for both short and longer term projects in these areas follow.

OBJECTIVE YIELD NONSAMPLING ERROR STUDIES

Goal: Monitor OY surveys for existing nonsampling errors, develop methods to detect new nonsampling errors, and evaluate methods to correct for the nonsampling errors.

Background: Recent studies, in addition to the various validation studies, included--

Sunflowers - enumerator versus combine-harvested yield;

Wheat - differences in lab and field counts;

Soybeans - plot biases in narrow row plantings, biases in field counts.

Work Plan: On-going and proposed studies include (project leader indicated in parenthesis)--

Wheat - (Vince Matthews) evaluation of unit differences before the use of a buffer zone was instituted to control some potential plot location biases, a draft is expected in early 1986. A draft of the evaluation of the new lab procedures and their effect on error reduction is expected by April 1986.

Corn - (Ron Steele) determining maturity category from destructive observations outside the unit, a draft is expected by May 1986. Yield differences between enumerator and farmer harvests are being analyzed by the Methods Staff.

Cotton - (Mike House) tagging effects on boll counts are being analyzed, a first draft is expected in early 1986.

Soybeans - (Bob Battaglia) effects of damage to plots from enumeration and differences between combine harvesting and hand harvesting, a draft is expected by mid 1986 for each study.

Sunflowers - a row space measurement study (Ron Fecso and Antoinette Tremblay) and yield changes between enumerator and farmer harvest (Vince Matthews with Doug Bond), a draft is expected by mid 1986 for each study.

The evaluation of unedited data and the use of prior months data will be evaluated (Fecso and Matthews) during the 1986 crop season if adequate ADP compatability can be arranged in an objective yield survey state.

OBJECTIVE YIELD SURVEY DESIGN AND ESTIMATION

Goal: Increase the precision and accuracy of OY survey estimates through design and estimator research and improve the hypothesis testing methods used with survey data.

Background: Undesirable properties have been identified in several estimation and analysis methods used with OY survey data. A recent study has shown that variance estimates are biased and improved estimators were proposed. For the same reason that variances are biased, current hypothesis testing methods have uncertain alpha levels.

Recent methodological changes include the use of a simplified soybean estimator; a reduction in the data collected in the soybean objective yield survey; and analysis of covariance as an improvement to the analysis of survey data.

Work Plans: Studies planned include --

1. Improved estimation of components of variance (segment, field, plot, data item within plot);

Improved estimation of components of variance (segment, field, plot, data item within plot);

2. Improved hypothesis tests and analysis techniques for complex survey data;

Improved hypothesis tests and analysis techniques for complex survey data;

3. Using resampling methods to improve estimators;

Using resampling methods to improve estimators;

4. Using multivariate distribution properties to develop data editing techniques; and

Using multivariate distribution properties to develop data editing techniques; and

5. Alternative gleaning units for corn objective yield.

Alternative gleaning units for corn objective yield.

Initial work on the first four of these studies will be conducted through an existing cooperative agreement with Iowa State University and Ron Fecso. Recommendations for further research in these areas are expected by mid 1986. Additional cooperative agreements are anticipated in order to complete the research. If recruiting efforts are successful, some projects may be done by Agency personnel.

The gleaning study is proposed by Ron Steele and Data Collection Branch. Collection of data is not expected before the 1987 crop year.

OBJECTIVE YIELD FORECASTING

Goal: Improve the yield forecasts made with objective yield survey data.

Background: Everyone wants better forecasts of yield and many people express a feeling that current forecasts are somewhat inadequate. Yet, explanations of the "inadequacies" have been vague and historically have been in the form of a suggestion to research some "piece" of the forecast model such as "maturity categories should be collapsed," or "we need a better weight model."

Recent emphasis has been to uncover and statistically define the problems with our current approach to forecasting. For example, fixed models have replaced stepwise models to remove unwarranted soybean forecast level changes and the inadvisability of combining data for different maturity categories in soybean forecast model development has been shown.

The forecasting problem can best be described as model misspecification. It appears that linear models are not an adequate way to describe the relationship between early season measurements and harvest data. One key paradox is the use of several years of historic data to develop the models. For model development, a long sequence of years is needed to account for trends in late season weather, but agronomic technology often changes over a few years dictating the need for a short sequence of years. Also, regression assumptions are violated by using clusters of data (years and pairs of plots) and a complex survey design rather than the independent and identically distributed observations upon which least squares parameter estimation is based.

Work Plan: First, more appropriate forms of models must be investigated. A promising candidate is a plot level production model rather than separate models for each of the several correlated components. Such a model will make a measure of plot forecast error available. This measure would be useful in the search for better predictor variables (which could include plant measurements, weather information or technology change data). Secondly, the corrections for the assumption violations must be sought.

The model improvements will be investigated by Agency personnel while the more theoretical problems will be addressed through cooperative agreements.

Currently, Vince Matthews plans to have a draft complete by mid 1986 which extends Tom Birkett's production estimator from corn to wheat. Weight per kernel data was collected in Kansas in 1985 for a study by Vince Matthews and Phil Doctor. A draft expected in mid 1986 will evaluate the use of the data in yield forecasting. Ron Steele will investigate technical deficiencies of the ratio model used for corn. A first draft is expected by October 1986. Steele is also examining the use of the corn maturity categories (draft is expected by October 1986) and evaluating the use of 2 additional years of data in corn models with publication expected by May 1986. Mike House will have results of alternative boll weight forecast methods by March 1986.

CROP MODEL ANALYSIS AND EVALUATION PROCEDURES

Goal: Develop, test and document variable selection and evaluation procedures employing design of experiments and response surface techniques to investigate the suitability and precision of component variables in any plant process model.

Background: During the past ten years SRS has been involved with the development and evaluation of plant process models for use in yield estimation and forecasting. Much of this work has resulted in little benefit to the agency. A viable model for either purpose has not been developed and candidates that may produce benefits in the near future do not appear on the horizon. There has been some progress in this area; however, scientists better understand the type of problem we are attempting to solve and at least two models have been developed utilizing the types of data collected in our objective yield program. Unfortunately, each of these models will require much additional development and will not be ready for any operational use for at least five years.

SRS has also learned from this involvement. One of the most valuable lessons is that the techniques needed to evaluate the important variables in a model have been developed or well defined. Variable evaluation is critical in applying these models since the user needs to be able to explain why the models react to the weather, soils, genetic and technological inputs. Further, the user also needs to know which of these inputs must be measured most precisely and which less precisely. These considerations are valuable in constructing the proper sampling and data collection procedures and the same information is valuable to identify why a model is not very realistic or useful.

In the future more and more people will be developing complex computer growth models. The analysis package developed in this goal will permit the developer or someone within SRS to determine the potential value of a model without the expense of a large scale data collection effort. Once a model is successfully analyzed and great promise is indicated then further data collection efforts and pilot studies are warranted.

Work Plan:	1. Conduct and document literature search.	Completed
	2. Draft proposal for variable selection specification, and analysis procedures.	Completed
	3. Develop and implement pilot software with Rolf Bargmann.	Completed
	4. Test pilot software on the Florida soybean model.	Completed
	5. Revise software and complete one final test of software.	2/86 - 6/86
	6. Document variable selection and evaluation procedures so that these techniques can be used by others.	7/86 - 9/86

Paul Cook is the technical coordinator.

PRODUCTION INPUT AND CROPPING PRACTICES

Goal: Collect production input and cropping practices information and evaluate its potential use for improving SRS corn and soybean yield forecasting and estimating procedures.

Background: Changing levels of production inputs and cropping practices affect crop yield potentials. Different economic conditions, agricultural policies and weather influence the applied levels of inputs in a given year and the adoption of new technologies. The purpose of this research is to measure the aggregate yield effects from various production technologies and to develop improved yield estimating procedures which take into account the technology, weather and soils interaction effects.

As part of the 1983 and 1984 corn and soybean objective yield survey, production input and cropping practices data were collected for sample fields in Illinois, Iowa and Missouri. Weather data from a nearby weather station and soils data from available county soil surveys were also collected for the sample fields. A similar data collection effort was done to supplement the 1985 survey.

Work Plans:

Activities	Time Table
1. Data Base Development	
a. Prepare 1983 data files (selected items from the OYS Forms A, B, D, H, and I summarized and merged with collected weather and soils data)	Completed
b. Prepare 1984 data files (all 1984 data files are complete.)	Completed
c. Build 1985 OYS data files	Dec. 1985-Mar 1986
d. Build 1985 weather and soils data files	Feb.-March 1986
2. Analysis of Potential Yield Relationships	
a. Combined analysis of data for all three years	Jan.-Apr 1986
b. Prepare draft of Staff Report	May 1986

The technical leader is Merritt Padgitt.

REVIEW NONPROBABILITY AND ISP YIELD INDICATION TO DEVELOP
REASONABLE FORECAST SERIES FOR THE NEW ISP SURVEY

Goal: Develop some bridge type relationships that will allow information from the historical nonprobability indications to be utilized in time series review for current ISP forecast indications.

Background: In the current implementation of the ISP survey more attention has been focused on the estimated items rather than the forecasted items. Nonprobability survey questions that generate the forecast indications are not requested in the ISP survey document. The new forecast indication thus stands alone without any historical performance information. Adjustments required in this new value are uncertain. As more States fall under the ISP program this question becomes more serious.

Work Plan: Work with the appropriate staff members in other branches to develop an evaluation program. Several possible approaches would be:

1. In place of crop condition, obtain an assessment of the crop using a five level or three level rating system, ie., (excellent, very good, good, bad, and very bad; or excellent, good, poor). This rating question would be asked of both probability and nonprobability respondents. It could be quantified by a discussion of percent of normal (average yield on this farm or in this locality). This type of data could be used in a weather trend model early in the growing season for crop forecasting. This type of variable could be used in a model directly or to select the number of standard deviations to depart from the trend forecast.
2. A matching analysis of early season yield projections to end of season yield for ISP respondents could be compared to a same type of analysis for farm report respondents. If the equations varied little in a given year or if the slopes of the equations were statistically the same then it may be possible to reconstruct mean historic yields from matched reports or even go as far as to use the mean locality yield. Level adjustments between indications would also need to be examined.
3. Additional ideas considered by the group would be explored. This effort would begin in the fall of 1985 and conclude in the spring of 1987. Some preliminary results would be provided every six months.
4. Data will be collected in several states during 1986.

Ben Klugh will provide analysis support.

DEVELOP OBJECTIVE PROCEDURES FOR CROP REPORTING BOARD ACTIONS

Goal: Develop objective procedures for crop reporting board actions that produce consistent and reliable forecasts and estimates and allow for evaluation of actions.

Background: As users of our data series becomes technically more competent, we have a responsibility to develop objective procedures to be utilized by the CRB to produce the best forecasts and estimates possible from our data. This problem requires expertise in both estimation theory, decision theory, and modeling. Procedures need to be developed to produce the estimates, as well as to monitor the current status of each indication.

- Work Plans:**
1. Using previously prepared data bases, examine the current proposed procedures for combining indications to produce an estimate. Prepare a first draft on these results by June 1986.
 2. Using data from crop and price surveys examine the advantages and disadvantages of multivariate estimators or regression estimators for acreage, yield and price estimates. Prepare a first draft by September 1986.
 3. Using previously prepared data bases, examine alternative approaches to the problem employing different estimation and modeling criteria. Prepare a first draft on these results by December 1986.
 4. Apply the best techniques developed in (1) to (3) to other indications and data series used by the Agency. These will be done on a priority basis determined by the Director, Chief, and section head.

Diane Willimack and Ben Klugh will provide the technical support.

DEVELOP A COMPOSITE FORECAST FROM SEVERAL STATISTICALLY SOUND WAYS
TO PREDICT YIELD FROM CORN AND COTTON OBJECTIVE YIELD DATA

Goal: Produce several reliable yield forecast methods for corn and cotton that can be combined using a composite forecast procedure, thus producing a more consistent and reliable yield forecast for board action.

Background: One problem with the current objective yield forecast models is that some of the assumptions employed to build the models are not satisfied. A solution to that problem would be to replace the current procedure with an alternative procedure or procedures for which the assumptions are satisfied. Fortunately, there are several alternatives that can be applied to current data that are as potentially reliable as each other. In forecasting when this dilemma occurs the solution is quite simple - composite estimation. Papers have been published that indicate that such composite forecasts perform better than any of the original individual forecasts. In addition, since these models will possess a strong statistical foundation, interpretation of why a change in forecast occurs will be easier to develop.

Work Plan: Work is currently underway to evaluate three alternative modeling techniques that require fewer assumptions or for which the required assumptions are better satisfied by the data. Two of these techniques have already been bootstrap tested against current procedures and both produce better forecasts. These first two methods are the state level production model and the probability compartment model. To complete this research will require another six to eight months so that a few more states can be evaluated and bootstrap testing completed.

The third method utilizes current modeling techniques but addresses the problem where the problem exists. How do we weight historic information to produce a forecast? In this method, sample level models are built using previous years for which model parameter estimates are the same. When a forecast is created there are several forecast models to choose from. Each model reflects some statistically different previous set of conditions. Forecasts from these models are weighted together using a weight function where the weights sum to one. These weights are our conditional belief that the forecast from one of these earlier models could apply in the current year. Weights can be developed from evaluation of weather and survey data. This method will require about one year to complete. While working in this area we will also be examining the use of some transformation algorithms that will optimize fit. This entire research project will be completed for corn by the summer of 1987 and be ready for implementation at that time. Cotton will be completed later.

Tom Birkett and Ben Klugh will be working on this effort.

DAILY WEATHER/SOIL MOISTURE YIELD MODELS

Goal: Develop, test and evaluate models for forecasting yield and/or weight per fruit for corn, soybeans and cotton. Such models would be based upon daily soil moisture balances, and other weather related data, in addition to objective survey measurements.

Background: The Yield Study Task Force's 1984 report recommended the use of supplemental information to increase the accuracy of the objective yield survey's forecast of weight per fruit. It was further recommended that the initial focus be on corn ear weight for Iowa and Illinois. The focus of the initial research into improved forecast procedures for soybeans and cotton will be in major producing states.

Work plan:

Activities	Status
Corn	
1. Creation of data bases (through 1984 crop season) required for model development and testing. These include:	
a. Yield and acreage, at the CRD level	Completed
b. Crop calendar	Iowa - Completed Illinois - March 15 Indiana - April 15 Missouri - April 15 Ohio - April 15
c. Daily weather - CRD level	Iowa, Illinois, Missouri and Nebraska available.
d. Corn OY survey averages and Board forecasts	Available
2. Model Development (Iowa)	March 1986
3. Model Testing and Evaluation	Illinois and Missouri, April and May 1986

Soybeans

Initial model development will be for Iowa and Mississippi and should start around July 1986.

Cotton

Initial model development will be for Mississippi and should start around March 1987.

Technical leaders are Fred Warren, Fatu Bigsby and Paul Cook.